

# **Propensity score matching in practice: The case of New Deal for Lone Parents**

**GSRU course  
January 18<sup>th</sup> 2008**

**Mike Daly,**

**Department of Work and Pensions**

# Introduction and outline

- Using example of NDLP
  - Groundbreaking at the time – and still is in some ways
  - Thoroughly reviewed later
- Overview of NDLP
- Why PSM?
- Outline of (impact) evaluation design
- Difficulties
- Overall reflections, lessons learned

# New Deal for Lone Parents

- Voluntary programme for lone parents claiming Income Support
  - Pilots began in 1997
  - National programme for new/repeat claims April 98
  - Extended to all claimants October 98
- Low take-up – around 5% at time of study
  - Planning began in 1999
- ‘Treatment’ consists largely of Personal Adviser service, with access to some training etc
- No set period of participation, no bar to repeat spells

# Choice of impact estimation approach

- National programme rules out an experiment
- Low take-up strongly suggests estimating impact on the treated, not the eligible
  - So diff-in-diff approach not feasible – even if suitable comparison group identified\*
- Regression unsuitable because of common support problem – 95% of eligible lone parents did not participate
- IV approach was suggested but unable to identify suitable instrument in advance

\*Although see Gregg & Harkness, 2003

# Design considerations

- In comparing participants and non-participants, selection bias is primary concern
- Recognise importance of (alleged) unobservables – motivation etc
- Some studies based on assumption that sufficiently rich work histories will reflect these
- We took alternative approach and measured them directly through a survey

# Collecting data on motivation

- Example questions:
  - A person must have a job to feel a full member of society
  - A woman and her family will benefit if she has a paid job
  - Having almost any job is better than being unemployed
  - It is less important for a woman to go out to work than for a man
  - Women with school-aged children should never work full-time
  - It is just wrong for a woman with children under five to go out to work
- As well as more familiar ones such as:
  - Whether actively looking for work
  - Preferred hours of work
- BUT... participation surely affects responses, so....

# Outline of design\*

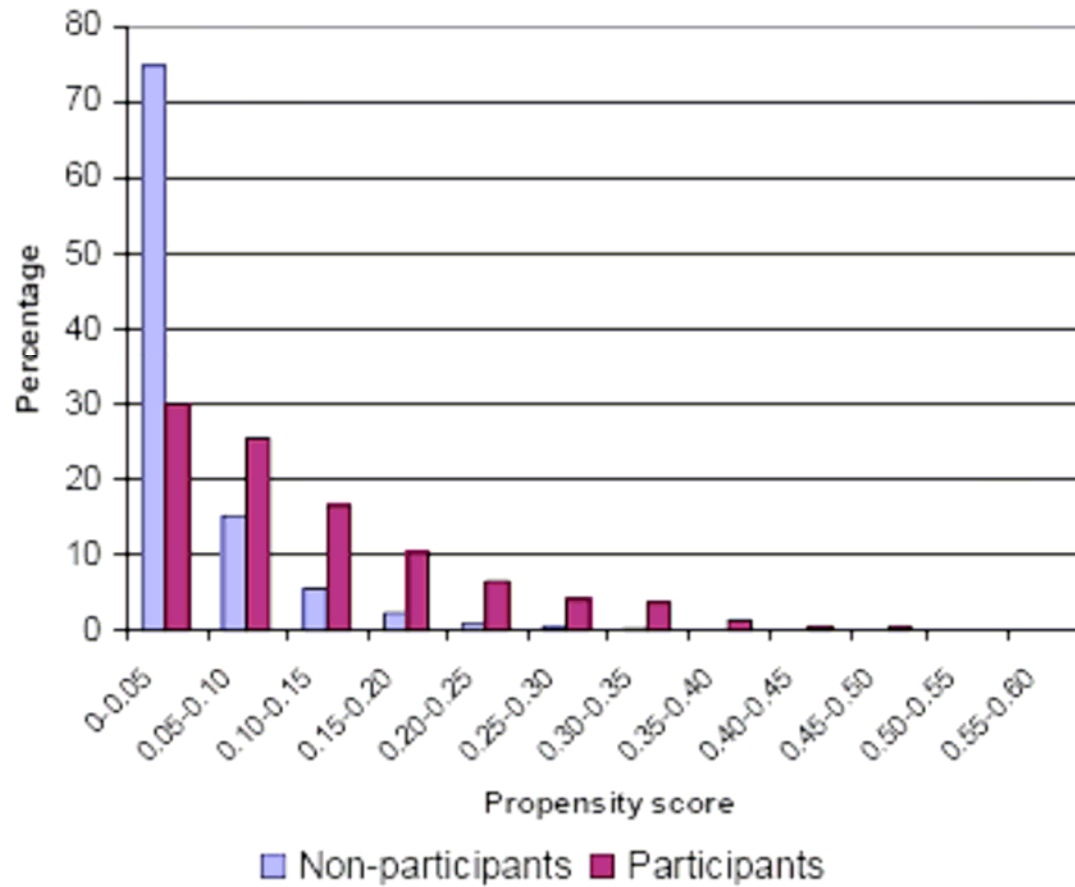
- Select a large sample (69,851) of eligible lone parents who have not yet participated in NDLP
- Send a postal questionnaire to all sample members to collect data on factors thought to affect the propensity to participate
- Track postal survey respondents using the NDLP evaluation database to identify new NDLP participants. This gives the participant sample
- Use the postal survey data in conjunction with administrative data to estimate a propensity score for each respondent
- Match each member of the participant sample to the non-participant with the nearest propensity score to create a matched sample of non-participants
- After a period of a few months carry out face to face interviews with all members of the participant and the matched comparison samples to collect data on outcomes
- Compare the rates of outcomes for participants and the matched sample of non-participants. The difference between the two is judged to be the estimate of additionality

\* Adapted from Lessof et al, 2003

# Some results

- Participants very different from the (unmatched) non-participants. E.g.
  - Unmatched – 23% of participants < 25; 48% had only one child, compared with 18 and 42% of non-participants
- Propensity score model did incorporate many items from postal survey – and in particular attitudinal data was significant
- Common support not a significant problem:

Figure 2.3 Distribution of propensity scores for participants and non-participants



NB Total sample is 42,273 so 0.025% is still 10 cases

\* from Lessof et al, 2003

# More results

- Programme had a substantial effect
- After 9 months, 49% of participants has entered work of 16+ hours, compared to 25% of matched non-participants
- Robust to a range of sensitivity analyses

# Difficulties/issues:

a) inherent to propensity score matching

- Lack of familiarity:
  - Takes some quantitative analysts - and policy makers with an analytical background - outside their comfort zone
  - Social researchers thinking it is harder than it is
  - Mistaken belief that it requires additional strong assumptions
  - (at that time) scarce expertise
- Many variations on the technique – sensitivity testing can be extensive

# Difficulties/issues:

b) common to many methods but exposed by propensity score matching

- Ultimately, cannot *prove* the Conditional Independence Assumption
- Requires very rich data to be plausible
- Difficulty of delayed and repeat participation
- Timetables of research constrain debate over results

# Difficulties/issues:

## C) peculiar to this study

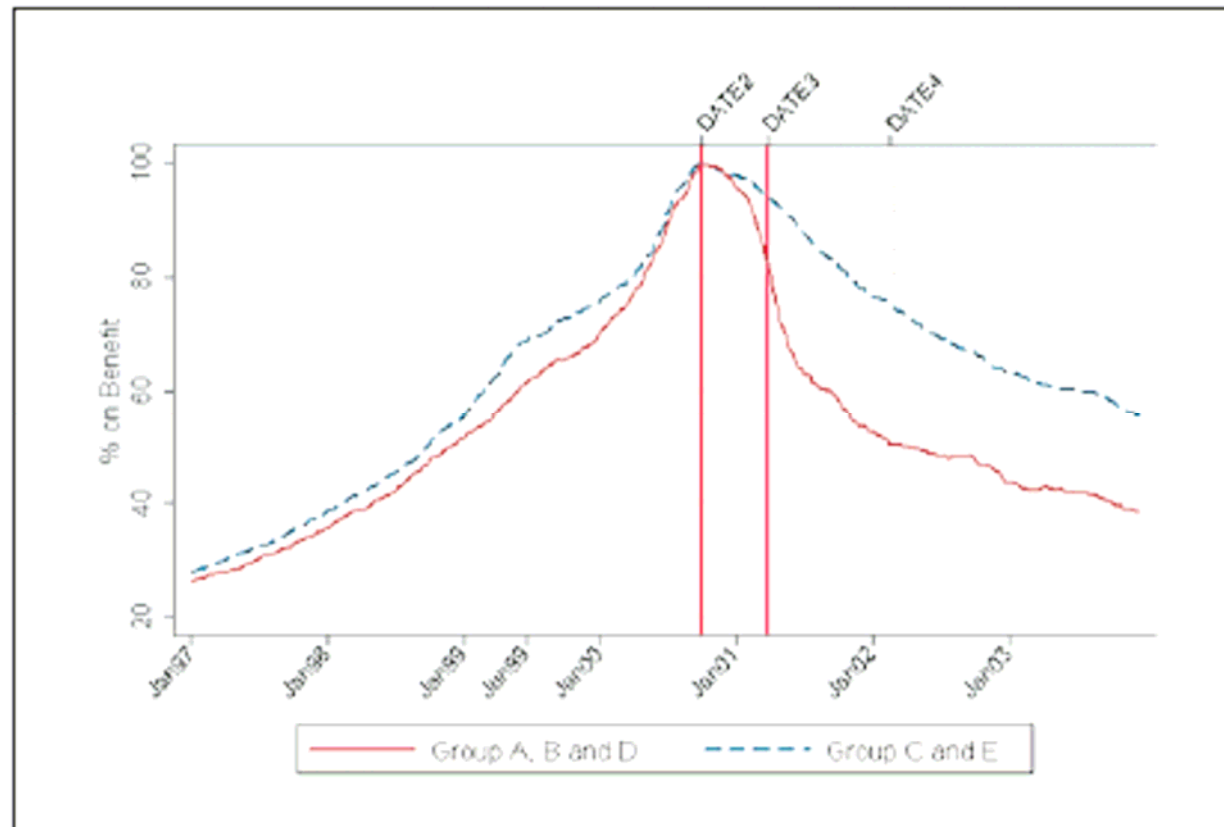
- Protracted – from conception in mid-99 to published report in March 2003 – and subsequent re-analysis in June 2006
- Multi-stage approach makes it difficult to deal with non-response
- Dealing with those who participated before the survey was issued, and non-participants who left benefit without participating
- How long can we assume surveyed attitudes remain unchanged?
- Difficulty of interpreting some responses – where the missing values seemed to be important
- Outcomes of matched non-participants not greatly different to the unmatched ones – arguably good news for the programme, but prompted some scepticism - particularly since the estimated impacts were so large
- And .... We made some mistakes

# Re-analysis of the data

- At time of publication, still more avenues we wanted to explore
- Particularly exploiting more recently available administrative data
- Commissioned some further analysis which found:

# Our matched sample was less matched than we thought

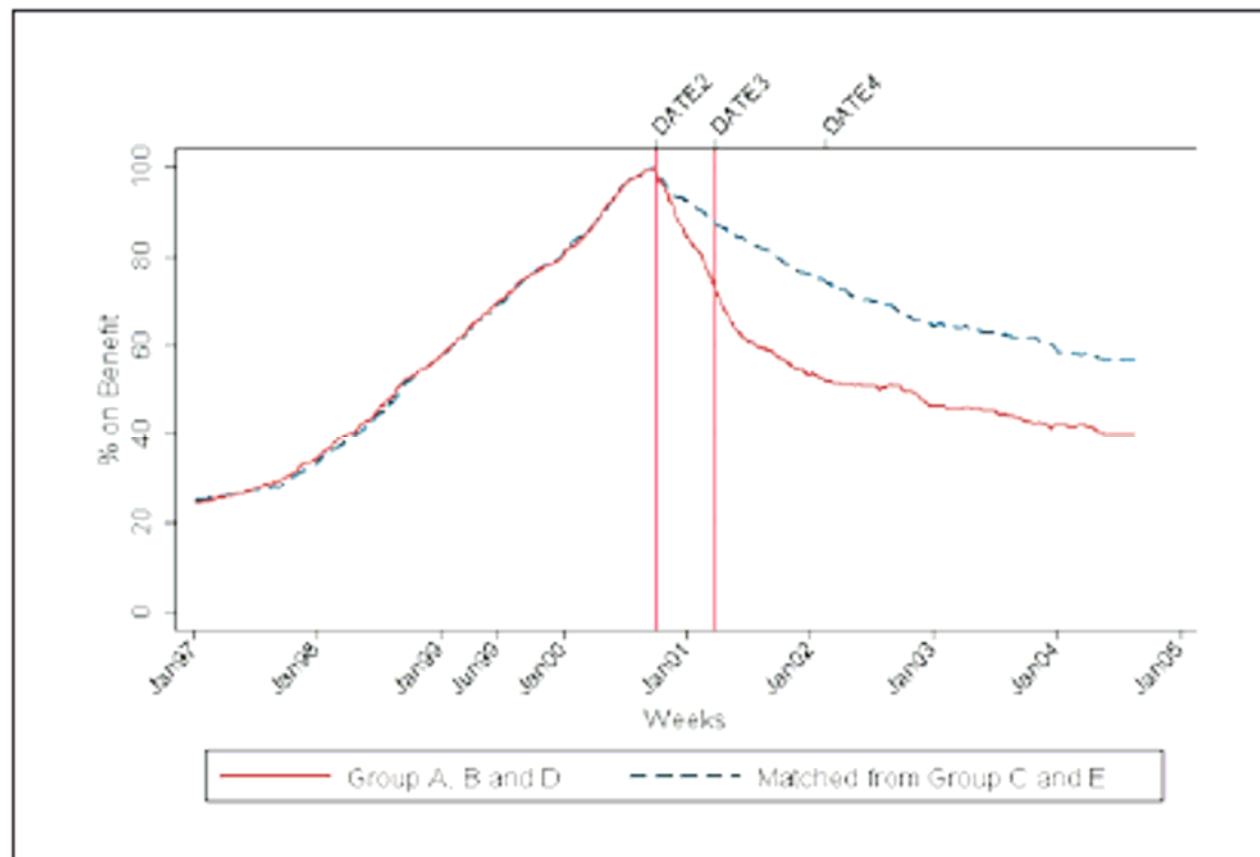
Figure 6.3 Weekly proportion of participants and non-participants on benefit – 1997-2003 (Sample III – Matched)



\* from Dolton et al, 2006

# Could match using admin data alone

Figure 6.6 Weekly proportion of participants and non-participants on benefit – 1997-2003 (Sample I – Specification 3)



\* from Dolton et al, 2006

## Although note that:

- Still cannot be certain that the samples match on unobserved characteristics
- Even if administrative data can achieve adequate matching in this case, cannot assume it always will
- And we certainly couldn't have assumed this in advance
- And it still shows a substantial positive impact – lasting around 4 years at least

# Reflections

- Was it the right approach?
  - Yes – it worked and produced robust results. Could not have anticipated that we could do without postal survey
- Was it worth it?
  - Yes – it was expensive (money and time) but is a major plank of DWP policy in this area
- Could we have done it better?
  - Undoubtedly! A number of lessons .....

# Lessons learned

- Be sure it's the right approach – ‘matching is not a magic bullet that will solve all evaluation problems’\*
- You need real expertise to get it right:
  - Project staff need to understand the approach
  - Ideally you need somebody internal with significant practical and theoretical knowledge
  - Research contractors similarly need most or all to be familiar, but to have some high-powered expertise
  - Consider consultancy if necessary
- Be realistic about time scales and budgets, and be convinced that it is value for money
- Be rigorous in documenting data and methods

\*Smith and Todd, 2000

# More lessons

- Good data is vital:
  - “... *the literature has spent relatively too much time worrying about estimator choice and relatively too little time worrying about data quality*” (Jeff Smith, 2000)
- Encourage alternative approaches:
  - Unlikely to be able to afford to fund duplicate studies
  - But be willing to make data available
- Encourage debate
  - Unlikely to get much academic interest in departmental publications – journal papers, conference participation are much better for this
- Look for evidence that matching has ‘worked’ in similar evaluations
  - e.g. as in NDDP

# Key references

## **New Deal for Lone Parents**

- New Deal for Lone Parents evaluation: findings from the quantitative survey ; Lessof et al; DWP report WAE 147 (2003)
- Evaluation of the New Deal for Lone Parents: Technical report for the quantitative survey; Phillips et al; DWP report WAE 146 (2003)
- The econometric evaluation of the New Deal for Lone Parents; Dolton et al; DWP report 356 (2006)
- Welfare reform and lone parents employment in the UK; Gregg and Harkness; CMPO working paper 03/072 (2003)

## **Work-based learning for Adults:**

- Work-based Learning for Adults: an evaluation of labour market effects; DWP report WAE 187 (2004)
- Evaluation of Work Based Learning for Adults: technical report; Taylor and Anderson; DWP report 189 (2004)
- The longer term outcomes of Work-Based Learning for Adults: Evidence from administrative data; Speckesser and Bewley; DWP report 390 (2006)

## **New Deal for Disabled People**

- Tests of non-experimental methods for evaluating the impact of New Deal for Disabled People; Orr, Bell and Kornfeld; DWP report WAE 198 (2004)
- Long-term impacts of the New Deal for Disabled People; Orr, Bell and Lam; DWP report 432 (2007)